

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 58 (2015) 461 – 469

Procedia
Computer Science

Second International Symposium on Computer Vision and the Internet (VisionNet'15)

A Comparative Study of Vision based Human Detection Techniques in People Counting Applications

Chakravartula Raghavachari, Aparna V, Chithira S, Vidhya Balasubramanian*

*Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham (University), Coimbatore*

Abstract

People counting has a wide range of applications in the context of pervasive systems. These applications range from efficient allocation of resources in smart buildings to handling emergency situations. There exist several vision based algorithms for people counting. Each algorithm performs differently in terms of efficiency, flexibility and accuracy for different indoor scenarios. Hence, evaluating these algorithms with respect to different application scenarios, environment conditions and camera orientations will provide a better choice for actual deployment. For this purpose, in our paper the most commonly implemented Frame Differencing, Circular Hough Transform and Histogram of Oriented Gradient based methods are evaluated with respect to different factors like camera orientation, lighting, occlusion etc. The performance of these algorithms under different scenarios demonstrates the need for more accurate and faster people counting algorithms.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

1. Introduction

People counting is a spatio-temporal function of human sensing, which gives the count of people in a particular area. Counting people is a useful task, which helps in understanding the flow of people in various places. The knowledge of density of people over an area would be helpful in handling emergency situations, efficient allocation of resources in the smart buildings etc. The constant movement of people, different age groups and body types makes people counting a challenging process. In addition, the presence of obstacles in indoor spaces etc., and varying lighting conditions make the process of accurately estimating the number of people in an area at given time very difficult.

1.1. Related Work

A wide range of technological solutions have been proposed for addressing the problem of people counting in both indoor and outdoor settings. People counting approaches can be mainly divided into two: instrumented and

* Corresponding author.

E-mail address: b_vidhya@cb.amrita.edu

uninstrumented approaches¹. Instrumented approaches can leverage wearable devices that give the individual count like RFID tags², WiFi tags¹⁰ etc. In many applications an instrumented approach would be infeasible since providing tags to people in public buildings is difficult and hence is not preferred.

On the other hand uninstrumented approaches rely on the external sensors like IR beam sensors, proximity sensors, laser sensors, temperature sensors, etc that are placed in the environment^{4,5,1}. IR and proximity sensors are widely used due to ease of use, but are not reliable when too many people arrive at the same time, and also produce many false positives due to other moving objects. Laser sensors work similarly to IR sensors⁵ and provide better accuracy at a higher cost. Thermal sensors are another option but there remain several challenges in using them for accurate counting. Cameras are a commonly used option and easy to instrument³. With the current progress in video analytics, counting using vision based approaches is getting more efficient.

Several camera based people counting techniques have been designed for different scenarios both in indoor and outdoor environments¹. One of the primary challenges in camera based approaches is the accurate detection of humans in different scenarios. Once the detection is done, counting is usually straightforward, therefore we outline the most popularly studied and implemented algorithms for human detection from a surveillance video.

Frame Differencing based background subtraction is the simplest approach that can be used for detecting moving objects²¹. In frame Differencing based people counting, the moving objects are detected and are classified as humans based on the bounding box parameters (viz length,width, area) that bounds moving objects. Gaussian Mixture Model^{12,13} is another background subtraction technique to detect moving objects. People Counting is achieved by normalizing the foreground objects extracted from the image and finding the edges using Gaussian Process Regression. In^{14,15}, Haar features are extracted from an image and they are classified as humans using a cascade classifier. In people counting Haar features are mainly used to detect and count the number of human faces.

In^{3,18}, human heads are detected and classified for counting using head classifier Statistically Effective Multi-scale Block Local Binary Pattern (SEMB-LBP). Here the head features are extracted and the count is obtained by counting the number of heads detected. Automatic people segmentation and counting based approach is proposed in¹⁶. Here K-means clustering with block wise background subtraction is used to segment individual person in the scene. Maximum number of clusters with suitable inter-cluster separation gives the number of people present in a scene.

Many recent techniques use Histogram of Oriented Gradient (HoG) based people counting, where histogram of gradient features of foreground objects are extracted and then classified using a classifier²³. This method can be applied for both inclined²³ and overhead cameras²⁴.

When overhead cameras are used, the features of the head and shoulder are usually considered. HoG features are extracted from the foreground and are classified using linear Support Vector Machine (SVM) classifier. In²⁰, Circular Hough transform (CHT) is used to detect human heads (as circles) in the image from an overhead camera in indoor space and counting is based on number of circles detected, and is shown to be faster for overhead cameras. However the performance of these approaches for scenarios which include both overhead and inclined cameras are yet to be evaluated.

1.2. Need for Comparative Study

As we have seen so far there are several vision based approaches for people counting. While many of these approaches work in detecting humans in different scenarios, when it comes to people counting as required by real-time pervasive systems there are several considerations:

- Firstly the technique must work irrespective of the camera orientation.
- The approach must be efficient so that real-time counting can be done.
- The solution must consider different scenarios, like entrances, large open spaces, lighting conditions, occlusions etc and provide accuracy in all these cases.

All the above camera based techniques work well in some cases, but their performance in terms of all these parameters namely accuracy, flexibility, efficiency, and robustness to camera orientation is not known, specially in indoor spaces. Therefore the commonly implemented camera based human detection approaches must be tested for these factors, so as to understand their suitability for people counting in indoor space.

Several studies have evaluated the different algorithms for people detection using vision based approaches. In^{9,25}, the authors have discussed about various methods involved in human sensing like presence, count, track etc. The authors have also discussed about various existing people counting approaches like instrumented and uninstrumented with common challenges that can occur while implementing these methods. However in the context of people counting there has not been a survey focusing on the specific application needs of people counting in indoor spaces to the best of our knowledge. In²³, existing techniques have been compared in the context of overhead cameras.

In this paper, a comparison is made between a few of the common vision based people counting algorithms with respect to different camera orientations and real time scenarios. For this purpose we have chosen Frame Differencing, CHT and HoG based people detection techniques. The first is the most common and cheapest solution. CHT is commonly used in the case of overhead cameras and HoG based approaches are still widely proposed (with some modifications to the basic HoG) and is flexible. Hence they form a broad range of approaches needed for our comparative study. The focus of this paper is to primarily understand how these fundamental approaches work in the context of the above mentioned requirements, and to gain new insights based on the performance of these algorithms over a wide variety of scenarios. We will be looking at some variations of HoG, but a full comparison of these variations is beyond scope of this paper.

These algorithms are evaluated in the context of counting people in an indoor area over at any given instance of time. We rigorously compare them under different lighting conditions, different crowd environments and camera orientations. We evaluate the accuracy, efficiency and response time of the different algorithms over these different scenarios. Based on our observations we suggest guidelines for deploying camera based people counting techniques.

The rest of the paper is structured as follows: Section 2 discusses the chosen algorithms for comparison. In Section 3 the experimental setup with results are discussed and we finally conclude in Section 4

2. Algorithms for Comparison

As discussed above, the task of detecting and counting humans who are continuously moving in different directions, varied illumination and backgrounds is challenging. Detection and counting of pedestrians can be performed by using various vision based algorithms, and as discussed earlier there is a need for a comparative study of algorithms in the context of people counting. In this section, we give an overview of the working of the chosen algorithms 1) Frame Differencing 2) Histogram of Oriented Gradients and 3) Circular Hough Transform.

2.1. Frame Differencing based people Counting

Frame differencing is a simple technique that can be implemented very easily for moving object detection²¹. This method extracts pixel wise differences of current frame from the previous frame to detect the moving objects. It is also popularly known as temporal differencing in which a background model (which is video frame at time t) is subtracted from a video frame at time $t + 1$. Inter frame difference method computes the absolute difference between the previous frame and current frame. The general operation performed in this method is given by:

$$Diff = |frame_{t+1} - frame_t| \quad (1)$$

This subtraction helps extract the foreground; if the absolute difference in pixel values for a given pixel position in both the images is greater than the threshold value T_s , then that pixel is considered as a part of the foreground. The thresholding condition is given by the following equation in which (i, j) represents the position of the pixel.

$$|frame_t(i, j) - frame_{t+1}(i, j)| > T_s \quad (2)$$

The detected moving objects are identified as humans by analysing the blobs, and their dimensions. People counting is performed by "bounding box with centroid" method²¹. For each blob, a bounding box is created and the number of such bounding boxes present over a scene at a particular time gives the people count. While this approach is easy to implement, there are false positives due to the possibility of similar dimension blobs that do not indicate humans.

2.2. Circular Hough Transform (CHT) based People Counting

This approach relies on detecting the circular shapes of human heads using overhead cameras. Hough transform is commonly used to determine the parameters of a circle whose equation can be represented as:

$$(x_i - a)^2 + (y_i - b)^2 = r^2. \quad (3)$$

(x_i, y_i) represents an arbitrary point of interest in the object's boundary, (a, b) represents the center of the circle, and r is the radius of the circle. For each point (x, y) on the original circle, a circle centered at (x_i, y_i) with radius R can be defined. The intersection point of all these circles in the parameter space yields the center of the desired circle in the original space. The algorithm can be similarly enhanced to determine multiple circles.

Once the heads are detected, the counting is done by counting the bounding boxes computed using a centroid method. The advantage of using Hough Transform for detection and counting is that it can handle the missing and occluded information when using an overhead camera. However it has to be adapted to perform with inclined cameras.

2.3. Histogram of Oriented Gradients based People Counting

HoG is an object detection technique, which uses distribution of intensity gradients or edge directions to define the shape of an object. It decomposes an image into cells, and for each cell, histogram of gradients are computed. Filtering is used to find the gradients in both horizontal and vertical directions. $[-1, 0, 1]$ and $[-1, 0, 1]^T$ are used as filters in the horizontal and vertical directions respectively. The combined histogram across all cells represents the descriptor. For better accuracy, normalization is done for all regions in the detection window by calculating a measure of local histogram over larger windows of fixed spatial regions. Once the features are detected, the humans are classified using a classifier. The training dataset includes annotated human images from different video streams. SVM classifiers are commonly used in such applications. Other classifiers like Adaboost which are cascading classifiers are also used for identifying and detecting humans. In this paper we study the performance of HoG along with an SVM classifier with both linear kernel and a non-linear kernel like RBF (Radial Basis Function).

3. Analysis of Different Algorithms

The previous sections explained the need for a comparative analysis of vision based human detection techniques and outlined the three algorithms as part of this study that address different challenges. This section explains the methodology of evaluation, the experimental setup and the results of these experiments.

We first discuss the system flow. The video feed from the camera is preprocessed, where frame extraction is done, and this processed feed is sent to the modules implementing the three different algorithms (Frame Differencing, CHT and HoG (with linear and RBF kernel)). Each of them individually detects the people in the feed, and sends this to the counting module. Here a simple counting algorithm is applied to determine the number of people in a region at a given instant of time.

3.1. Experimental Strategy

Once the different algorithms detect the humans, we need to evaluate their performance. For this we need to run the algorithms under different contexts as discussed in Section 1. The different scenarios for which the algorithms are evaluated are as follows:

- Camera Orientation: The camera can be either in an overhead position or in an inclined position. Both orientations are used for people counting in different contexts, and hence must be analysed.
- Density of the people: The human detection accuracy depends on the density of people in the coverage region. It can be low, medium or high. In different times of day and situations, the density can vary, and a people counting algorithm must work for different densities of people.
- Lighting conditions: Lighting affects vision based object detection algorithms. People counting algorithms must be robust to low lighting conditions and possible glare.

- Presence of occlusion: Humans can be occluded by other humans or objects in the environment. The effectiveness of any algorithms in the presence of such occlusions would provide an indication of suitability in real life scenarios.

To perform the experiments, we have chosen a lab environment, which is 210X110cm in dimension where there is a continuous flow of people. A camera is placed facing the entrance to the lab. The camera is positioned in both overhead and inclined orientations (approx. 60 degrees) covering the region of interest. Our setup for both the orientations are shown in figure 1 and figure 2.



Fig. 1. Overhead Position



Fig. 2. Inclined Position

Our objective is to find the number of people that are crossing the entrance of the lab (in both directions). An OpenCV based system is developed for implementing the algorithms given in Section 2. All the algorithms were implemented on a desktop machine with an Intel Core: i3 processor and 4.00GB RAM. The data used for testing the algorithms include video feeds from the camera that was collected over a period of 2-3 weeks under varied lighting conditions and various densities of people flow. The people count given by the individual algorithms is compared with actual count (done manually) to determine the accuracy. For each of the different scenarios multiple cases are generated.

The following are the different metrics that are used for evaluating the selected algorithms.

Detection Rate is the ratio of humans detected correctly to the total number of humans detected in the scene. It is defined as:

$$\text{DetectionRate} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \quad (4)$$

Recall is the ratio of number of people correctly detected to the actual number of people in the scene. It is defined as follows:

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \quad (5)$$

Accuracy is the ratio of true values to the sum of all the possible values²³. It is defined as follows:

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegatives}}{\text{TruePositive} + \text{FalsePositives} + \text{TrueNegatives} + \text{FalseNegative}} \quad (6)$$

Response Time is the time taken by the each algorithm to perform people counting.

3.2. Experimental Results

In this section we evaluate the selected vision based algorithms and some of their variants with respect to the categories explained in Section 3.1 and discuss the results and observations.

3.2.1. Orientation of the camera

The first set of experiments evaluates the impact of the camera orientation on people detection accuracy. For this purposes, the cameras have been placed both at overhead and inclined positions. Keeping orientation constant, cases spanning the other scenarios have been collected and the performance of the different strategies is compared. The Table shows the results obtained under this category. From Table 1 it is evident that the CHT performs best

Table 1. Performance under different camera orientations

		Overhead			Inclined		
		Detection Rate	Recall	Accuracy	Detection Rate	Recall	Accuracy
Frame Differencing		0.878	0.496	0.4527	0.902	0.3417	0.3195
Circular Hough Transform		0.853	0.84	0.736	0.534	0.646	0.418
HoG +	SVM (Linear)	0.790	0.911	0.736	0.8615	0.8375	0.7647
	SVM (RBF)	0.9196	0.8604	0.7978	0.9291	0.8792	0.8071

in overhead position of the camera. This is because the occlusion caused by other person is avoided in an overhead camera. On the other hand, HoG performs better in the inclined orientation. In inclined orientation, the complete view of people enables the extraction of human features well. Hence, HoG performs better for inclined orientation. When a SVM classifier with RBF kernel is used, the accuracy is improved a bit for both orientations. It must be noted that the HoG is only generally trained for human detection. If the training is specific to camera orientation, then it may perform more accurately. Frame differencing is the worst in terms of accuracy, and fares poorly in inclined orientation due to occlusions.

3.2.2. Various densities of people

Table 2 shows the results obtained for the people detection algorithms different densities of people in the region of interest. We consider different cases with low (1-2 people/frame), medium (3-4 people/frame) and high(5 people/frame) densities of people. From the table we can infer that frame differencing has the lowest accuracy compared to the other two algorithms. It is because when density is high i.e. when people are moving very close to each other it counts them as one. HoG performs best since it works consistently irrespective of camera orientation. Here again, the use of RBF kernel slightly improves the accuracy.

Table 2. Impact of density of people

		Low (1-2)			Moderate (3-4)			High (5)		
		Detection Rate	Recall	Accuracy	Detection Rate	Recall	Accuracy	Detection Rate	Recall	Accuracy
Frame Differencing		0.8108	0.6708	0.5646	0.8450	0.4540	0.480	0.775	0.275	0.250
Circular Hough Transform		0.6583	0.795	0.677	0.7362	0.8017	0.671	0.8369	0.9	0.6348
HoG +	SVM (Linear)	0.6842	0.775	0.7625	0.6518	0.763	0.724	0.9	0.8	0.6935
	SVM (RBF)	0.9167	0.8316	0.7719	0.8743	0.8127	0.7581	0.8732	0.8094	0.7358

Figure 3 demonstrates this clearly, i.e, Frame Differencing here outputs count as one even though there are five people who are very close to each other. Figure 4 shows HoG for an inclined camera with a high density of people, demonstrating the better performance of this method.

3.2.3. Different lighting conditions

Next, we consider the different lighting conditions, i.e., the performance of the algorithms during the day time and night time. Table 3 shows how these algorithms perform under the two different modes. It is obvious that the accuracy for all the algorithms in day light is less when compared to night mode because of the presence of shadows in day light condition. Interestingly the linear kernel works better in the day light conditions and has better recall. This is also reiterated in the presence of occlusions as we will see next.

3.2.4. Presence of occlusion

Under this category, the algorithms are evaluated for both occluded and non occluded case. Table 4 shows the results obtained. It is obvious that the accuracy of each algorithm is higher in the non-occluded case when compared

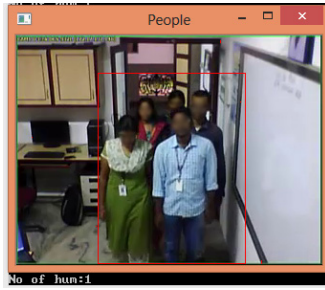


Fig. 3. Frame Differencing- High Density

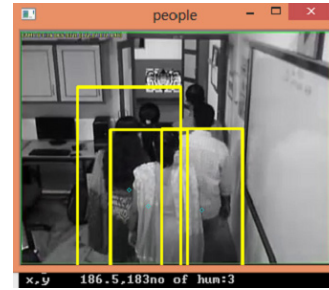


Fig. 4. HoG-High Density

to that of occluded case. In the presence of occlusions, HoG has the highest accuracy than the other two since the HoG is robust in capturing human features under different conditions. As mentioned before a linear kernel performs better in this case.

Table 3. Performance under different lighting conditions

		Night			Day		
		Detection Rate	Recall	Accuracy	Detection Rate	Recall	Accuracy
Frame Differencing		0.8491	0.4356	0.411	0.825	0.392	0.3455
Circular Hough Transform		0.685	0.7689	0.599	0.708	0.733	0.542
HoG +	SVM (Linear)	0.8013	0.864	0.776	0.8677	0.891	0.7090
	SVM (RBF)	0.8762	0.783	0.8011	0.8592	0.7264	0.7865

Table 4. Impact of occlusion

		Occlusion			No Occlusion		
		Detection Rate	Recall	Accuracy	Detection Rate	Recall	Accuracy
Frame Differencing		0.8181	0.338	0.3095	0.9208	0.6619	0.6165
Circular Hough Transform		0.7105	0.7284	0.5850	0.6440	0.7869	0.6527
HoG +	SVM (Linear)	0.8350	0.8916	0.7139	0.8414	0.8226	0.7448
	SVM (RBF)	0.8574	0.8257	0.7962	0.8729	0.7906	0.8199

Figure 5 demonstrates the robustness of HoG even when occlusion is present. This capacity of HoG features to detect humans even in the presence of occlusion makes it a desirable algorithm for people counting applications. Figure 6 demonstrates a case where CHT fails. Here it fails to detect a person whose head is mildly occluded, demonstrating that head detection is not a fail-safe approach.

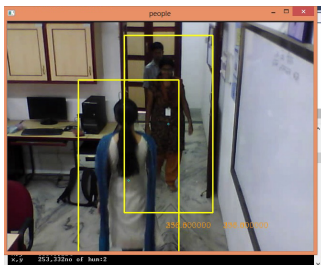


Fig. 5. Impact of Occlusion on HoG

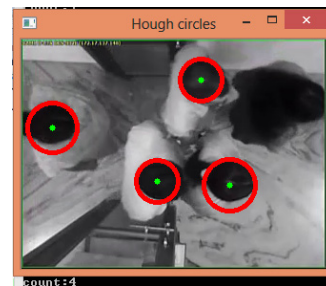


Fig. 6. Impact of Occlusion on Hough

Finally we analyze the overall accuracy and response of these algorithms. The metrics are aggregated based on all the above results are shown in Table 5.

Table 5. Overall Accuracy and Response Time

	Accuracy	Response Time (approx.)
Frame Differencing	0.4165	14 seconds
Circular Hough Transform	0.6060	46 seconds
HoG + SVM (Linear)	0.7404	112 seconds
SVM (RBF)	0.7860	87 seconds

We can see that HoG has the highest accuracy followed by the CHT method. The CHT works best for overhead cameras but fails in inclined positions. However it is much more efficient in terms of response time. HoG is robust to different scenarios, but has poor response time due to the classification requirement. The use of a RBF kernel results in better accuracy and better response time than the linear kernel. For real-time purposes faster techniques must be developed. The overall accuracy is less for all the algorithms when compared to that of^{19,20,21} since the data set which we used is more diverse, caters to both camera orientations and has more cases of occlusion.

4. Conclusion

In this paper we have compared three camera based approaches for human detection in the context of people counting. We have analyzed their performance in the context of different camera orientations, density of people, lighting conditions and occlusions. After analyzing them under different scenarios, we can have observed that the Histogram of Gradients based approach is most robust, even when it isn't trained specifically for specific camera orientations. Its performance in the presence of occlusions is very good and hence makes it a good candidate for such applications. Additionally using an RBF kernel helps improve the performance than using a linear kernel. However HoG based approach with SVM suffers from poor response time, and has to be addressed. The frame differencing method is easy to implement, and fast, but has poor accuracy. Hough, performs well when the camera is in the overhead position, but fails in the other orientation. This study shows the techniques that use classification over features in representing humans works effectively and more accurately. However it also demonstrate the needs for faster human detection algorithms, that do not require complex training and classification, so that they can be used in real-time systems. Our future work involves taking lessons from this study and developing robust and real-time algorithms for people counting.

Acknowledgements

This work has been funded in part by Department of Science and Technology, Government of India, grant DyNo. 100/IFD/2764/2012-2013.

References

1. Thiago Teixeira, Gershon Dublon, Andreas Savvides, "A Survey of Human-Sensing: Methods for Detecting Presence, Count, Location, Track, and Identity," *ENALAB technical report*, 2010.
2. Suguna P Subramanian, Jrgen Sommer, Stephen Schmitt and Wolfgang Rosenstiel, "RIL Reliable RFID based Indoor Localization for Pedestrians," *IEEE International Conference on Software, Telecommunications and Computer Networks*, pp. 218-222, 2008.
3. Zebin Cai, Zhu Liang Yu, Hao Liu and Ke Zhang, "Counting People in Crowded Scenes by Video Analyzing," *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1841-1845, 2014.
4. Kazuhiko Hashimoto, Katsuya Morinaka, Nobuyuki Yoshiike, Chjihiro Kawaguchi and Satoshi Matsueda, "People Count System Using Multi-Sensing Application," *IEEE International Conference on Solid State Sensors and Actuators*, vol. 2, pp. 1291-1294, 1997.
5. Huijing Zhao and Ryosuke Shibasaki, "A Novel System for Tracking Pedestrians Using Multiple Single-Row Laser-Range Scanners," *IEEE Transactions on Systems, Man, And Cybernetics Part A: Systems And Humans*, vol. 35, no. 2, pp. 283-291, 2005.
6. Andre Treptow, Grzegorz Cielniak and Tom Duckett, "Real-time People Tracking for Mobile Robots using Thermal Vision," *Robotics and Autonomous Systems, Elsevier*, vol. 54, no.9, pp. 729-739, 2006.

7. Ya-Li Hou, and Grantham K. H. Pang, "People Counting and Human Detection in a Challenging Situation," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems And Humans*, vol. 41, no. 1, pp. 24-33, 2011.
8. Zheng Ma and Antoni B. Chan, "Crossing the Line: Crowd Counting by Integer Programming with Local Features," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2539-2546, 2013.
9. Thomas B. Moeslund, Erik Granum, "A Survey of Computer Vision-Based Human Motion Capture," *IEEE Transactions on Computer Vision and Image Understanding*, vol. 81, Issue 3, pp. 231268, 2001.
10. Wei Xi, Jizhong Zhao, Xiang-Yang Li, Kun Zhao, Shaojie Tang, Xue Liu and Zhiping Jiang, "Electronic Frog Eye: Counting Crowd Using WiFi," *IEEE INFOCOM-IEEE Conference on Computer Communications*, pp.361-369. 2014.
11. Osama Masoud and Nikolaos P.Papanikolopoulos, "A Novel Method for Tracking and Counting Pedestrians in Real-Time Using a Single Camera," *IEEE Transactions on Vehicular Technology*, vol.50, no.5, pp. 1267-1278, 2001.
12. Chan Antoni B and Vasconcelos Nuno "Counting People with Low-level Features and Bayesian Regression," *IEEE Transactions on Image Processing*, vol.21, pp. 2160-2177, 2012.
13. Huazhong Xu, Pei Lv and Lei Meng "A People Counting System based on Head-shoulder Detection and Tracking in Surveillance Video," *IEEE International Conference on Computer Design and Applications (ICCD)*, vol.1, pp.V1-394, 2010.
14. Chen Duan-Sheng and Liu Zheng-Kai "Generalized Haar-like features for fast face detection," *International Conference on Machine Learning and Cybernetics*, vol.4, pp.2131-2135, 2007.
15. Geismann Philip and Schneider Georg "A two-staged approach to vision-based pedestrian recognition using Haar and HOG features," *IEEE Symposium on Intelligent Vehicles*, pp.554-559 2008.
16. Antic Borislav, Letic Dragan Culibrk Dubravko and Crnojevic Vladimir "K-means based segmentation for real-time zenithal people counting," *IEEE International Conference on Image Processing (ICIP)*, pp.2565-2568, 2009.
17. Kah-Kay Sung and Tomaso Poggio "Example-Based Learning for View-Based Human Face Detection," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, pp.39-51, 1998.
18. Chengbin Zeng and Huadong Ma "Robust Head-shoulder Detection by PCA-Based Multilevel HOG-LBP Detector for People Counting," *IEEE International Conference on Pattern Recognition*, pp.2069-2072, 2010.
19. Navneet Dalal and Bill Triggs "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol.1, pp. 886-893, 2005.
20. Jingyu Liu, Jiazheng Liu and Mengyang Zhang, "A Detection And Tracking Based Method For Real-time People Counting," *Chinese Automation Congress (CAC)*, IEEE, pp. 470-473, 2013.
21. Kin-Yi Yam, Wan-Chi Siu, Ngai-Fong Law and Chok-Ki Chan, "Effective bi-directional people flow counting for real time surveillance system," *IEEE International Conference on Consumer Electronics (ICCE)*, vol. 11, pp. 863-864, 2011.
22. Liping CHEN, Huibin WU, Shuguang ZHAO and Jiong GU, "Head-shoulder Detection Using Joint HOG features for People Counting and Video Surveillance in Library," *IEEE Workshop on Electronics, Computer and Applications*, pp. 429-432, 2014.
23. Bin Li, Jian Zhang, Zheng Zhang and Yong Xu, "A People Counting Method Based on Head Detection and Tracking," *IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 136-141, 2014.
24. Jorge Garca, Alfredo Gardel, Ignacio Bravo, Jos Luis Lzaro, Miguel Martnez, and David Rodrguez, "Directional People Counter Based on Head Tracking," *IEEE Transactions On Industrial Electronics*, vol. 60, no. 9, pp. 3991-4000, 2013.
25. Thomas B. Moeslund, Adrian Hilton, and Volker Krger, "A survey of Advances in Vision-based Human Motion Capture and Analysis," *A survey of Advances in Vision-based Human Motion Capture and Analysis*, vol. 104, no.2, pp.90126, 2006.